University of Southern Denmark
IMADA
Arthur Zimek
Jonatan Møller Gøttcke, Jonas Herskind Sejr

DM566/DM868/DM870: Data Mining and Machine Learning

Spring term 2019

Exercise 9: Probability, Conditional Probability, Bayes' Theorem, Naïve Bayes

Exercise 9-1 Events and Sample Spaces

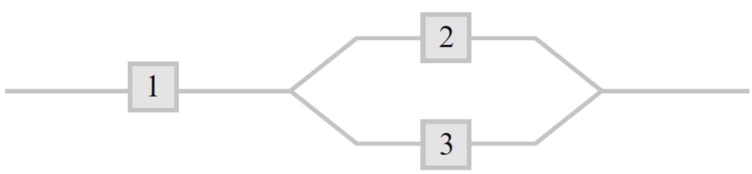
(a) We have a system of several fuses. We can examine each single fuse to see whether it is defective. The sample space for this experiment can be abbreviated as $\Omega = \{N, D\}$, where N represents not defective, D represents defective.

If we examine three fuses in sequence and note the result of each examination, what is the sample space Ω?

(b) As an experiment, we observe the number of pumps in use at a six-pump gas-station, so simple events are the numbers 0-6 (pumps in use). Given the events $A = \{0, 1, 2, 3, 4\}$, $B = \{3, 4, 5, 6\}$, and $C = \{1, 3, 5\}$, which simple events are contained in

(i) $A \cup B$?
(ii) $A \cup C$?
(iii) $A \cap B$?
(iv) $A \cap C$?
(v) $\overline{A}$?
(vi) $\overline{A \cup C}$?

(c) Three components are connected to form a system as shown in this diagram:



Because the components in the 2-3 subsystem are connected in parallel, that subsystem will function if at least one of the two individual components functions. For the entire system to function, component 1 must function and so must the 2-3 subsystem. The experiment consists of determining the condition of each component (S (success) for a functioning component and F (failure) for a non-functioning component).

(i) What outcomes are contained in the event A that exactly two out of the three components function?
(ii) What outcomes are contained in the event B that at least two of the components function?
(iii) What outcomes are contained in the event C that the system functions?
(iv) List the outcomes in $\overline{C}$, $A \cap C$, $A \cup C$, $B \cup C$, and $B \cap C$.

Exercise 9-2 Conditional Probability

Suppose that of all individuals buying a certain digital camera, 60% include an optional memory card in their purchase, 40% include an extra battery, and 30% include both a card and battery. Consider randomly selecting a buyer and let $A = \{\text{memory card purchased}\}$ and $B = \{\text{battery purchased}\}$.

Then $\Pr(A) = 0.6$, $\Pr(B) = 0.4$, and $\Pr(\text{both purchased}) = \Pr(A \cap B) = 0.3$.

(a) Given that the selected individual purchased an extra battery, what is the probability that an optional card was also purchased?

(b) Given that the selected individual purchased a memory card, what is the probability that an optional extra battery was also purchased?

Exercise 9-3 Bayes' Theorem

Only 1 in 1000 adults is afflicted with a rare disease for which a diagnostic test has been developed. The test is such that when an individual actually has the disease, a positive result will occur 99% of the time, whereas an individual without the disease will show a positive test result only 2% of the time.

If a randomly selected individual is tested and the result is positive, what is the probability that the individual has the disease?

Exercise 9-4 Naïve Bayes

The skiing season is open. To reliably decide when to go skiing and when not, you could use a classifier such as Naïve Bayes. The classifier will be trained with your observations from the last year. Your notes include the following attributes:
The weather: The attribute `weather` can have the following three values: `sunny`, `rainy`, and `snow`.
The snow level: The attribute `snow level` can have the following two values: ≥ 50 (There are at least 50 cm of snow) and < 50 (There are less than 50 cm of snow).

Assume you went skiing 8 times during the previous year. Here is the table with your decisions:

weather	**snow level**	**ski ?**
sunny	< 50	no
rainy	< 50	no
rainy	≥ 50	no
snow	≥ 50	yes
snow	< 50	no
sunny	≥ 50	yes
snow	≥ 50	yes
rainy	< 50	yes

(a) Compute the *a priori* probabilities for both classes `ski = yes` and `ski = no` (on the training set)!

(b) Compute the distribution of the conditional probabilities for the two classes for each attribute.

(c) Decide for the following weather and snow conditions, whether to go skiing or not! Use the Naïve Bayes classifier as trained in the previous steps for your decision.

	weather	**snow level**
day A	sunny	≥ 50
day B	rainy	< 50
day C	snow	< 50